

Tilburg University

TraMOOC: Translation for Massive Open Online Courses

Kordoni, Valia; Cholakov, Kostadin; Egg, Markus; Way, Andy; Birch, Lexi; Kermanidis, Katia; Sosoni, Vilelmini; Tsoumakos, Dimitrios; van den Bosch, Antal; Hendrickx, I.H.E.; Papadopoulos, Michael; Georgakopoulou, Panayota; Gialama, Maria; van Zaanen, Menno; Buliga, Ioana; Jermol, Mitja; Orlic, Davor

Published in:

Proceedings of the 18th Annual Conference of the European Association for Machine Translation (EAMT2015)

Publication date:

2015

Document Version

Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):

Kordoni, V., Cholakov, K., Egg, M., Way, A., Birch, L., Kermanidis, K., Sosoni, V., Tsoumakos, D., van den Bosch, A., Hendrickx, I. H. E., Papadopoulos, M., Georgakopoulou, P., Gialama, M., van Zaanen, M., Buliga, I., Jermol, M., & Orlic, D. (2015). TraMOOC: Translation for Massive Open Online Courses. In I. Durgar El - Kahlout, M. Özkan, F. Sánchez - Martínez, G. Ramírez - Sánchez, F. Hollowood, & A. Way (Eds.), *Proceedings of the 18th Annual Conference of the European Association for Machine Translation (EAMT2015)* (pp. 217). EAMT.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



TraMOOC: Translation for Massive Open Online Courses

Funding agency: The European Commission

Funding call identification: H2020-ICT-2014-1 - ICT-17-2014

Type of project: Innovation Action

Project ID number: 644333

<http://www.tramooc.eu>

| List of partners |
|---|
| Humboldt Universität zu Berlin (UBER), Germany (coordinator) |
| Dublin City University (DCU), Ireland |
| The University of Edinburgh (UEDIN), UK |
| Ionian University (IURC), Greece |
| Stichting Katholieke Universiteit (Radboud University & Radboud UMC), The Netherlands |
| EASN Technology Innovation Services BVBA (EASN TIS), Belgium |
| Deluxe Media Europe Ltd (Deluxe Media Europe Ltd), UK |
| Stichting Katholieke Universiteit Brabant Universiteit van Tilburg, The Netherlands |
| IVERSITY GMBH (iversity.org), Germany |
| KNOWLEDGE 4 ALL FOUNDATION (K4A), UK |

Project duration: February 2015 — January 2018

Summary

Massive open online courses have been growing rapidly in size and impact. TraMOOC aims at developing high-quality translation of all types of text genre included in MOOCs from English into eleven European and BRIC languages (DE, IT, PT, EL, DU, CS, BG, CR, PL, RU, ZH) that are hard to translate into and have weak MT support. Phrase-based and syntax-based SMT models will be developed for addressing language diversity and supporting the language-independent nature of the methodology. For a high quality, automatic translation approach and for adding value to existing infrastructure, extensive and advanced bootstrapping of new resources will be performed. An innovative multi-modal automatic and human evaluation schema will further ensure translation quality. For human evaluation, an innovative, strict-access control, time- and cost-efficient crowdsourcing setup will be used. Translation experts, domain experts and end users will also be involved. Separate task mining applications will be employed for implicit translation evaluation: (i) topic detection will be applied to source and translated texts and the resulting entity lists will be compared, leading to further qualitative and quantitative translation evaluation results; (ii) sentiment analysis performed on MOOC users' blog posts will reveal end user opinion/evaluation regarding translation quality. Results will be combined into a feedback vector and used to refine parallel data and retrain translation models towards a more accurate second phase translation output. The project results will be showcased and tested on the Iversity MOOC platform and on the VideoLectures.net digital video lecture library.